GAATTCCAGTCTGCTGGCTTCC) .CTGCAGCACAGCACACTCCCTTTGGCAAGGACCTGAGACCCTTGTGC. .GTC

```
                                                              1
                                                              met ile phe pro trp lys [cys]
AAGAGGCTCAATGGGCTGCAGAAGAACTAGAGAAGGACCAAGCAAAGCC    ATG ATA TTT CCA TGG AAA [TGT]

         10                                          20              Signal Sequence
gln ser thr gln arg asp leu trp asn ile phe lys [leu trp gly trp thr met leu [cys]
CAG AGC ACC CAG AGG GAC TTA TGG AAC ATC TTC AAG  TTG TGG GGG TGG ACA ATG CTC [TGT]

       30                            ┌─ Probable N-Terminus
[cys]asp phe leu ala]his his gly thr tyr [cys]trp thr tyr his tyr ser glu lys pro
[TGT]GAT TTC CTG GCA]CAT CAT GGA ACC TAC [TGC]TGG ACT TAC CAT TAT TCT GAA AAA CCC

       50                                        60
met asn trp gln arg ala arg arg phe [cys]arg asp[asn tyr thr]asp leu val ala ile
ATG AAC TGG CAA AGG GCT AGA AGA TTC [TGC]CGA GAC[AAT TAC ACA]GAT TTA GTT GCC ATA

       70                                80
gln asn lys ala glu ile glu tyr leu glu lys thr leu pro phe ser arg ser tyr tyr
CAA AAC AAG GCG GAA ATT GAG TAT CTG GAG AAG ACT CTG CCC TTC AGT CGT TCT TAC TAC

       90                              100
trp ile gly ile arg lys ile gly gly ile trp thr trp val gly thr[asn lys ser]leu
TGG ATA GGA ATC CGG AAG ATA GGA GGA ATA TGG ACG TGG GTG GGA ACC[AAC AAA TCT]CTC

      110                             120
thr glu glu ala glu asn trp gly asp gly glu pro asn asn lys lys asn lys glu asp
ACT GAA GAA GCA GAG AAC TGG GGA GAT GGT GAG CCC AAC AAC AAG AAG AAC AAG GAG GAC

      130                             140
[cys]val glu ile tyr ile lys arg asn lys asp ala gly lys trp asn asp asp ala[cys]
[TGC]GTG GAG ATC TAT ATC AAG AGA AAC AAA GAT GCA GGC AAA TGG AAC GAT GAC GCC[TGC]

      150
his lys leu lys ala ala leu[cys]tyr thr ala ser[cys]gln pro trp ser[cys]ser gly
CAC AAA CTA AAG GCA GCC CTC[TGT]TAC ACA GCT TCT[TGC]CAG CCC TGG TCA[TGC]AGT GGC

      170                           180
his gly glu[cys]val glu ile ile asn[asn his thr]cys]asn[cys]asp val gly tyr tyr
CAT GGA GAA[TGT]GTA GAA ATC ATC AAT[AAT CAC ACC]TGC]AAC[TGT]GAT GTG GGG TAC TAT

      190                             200
gly pro gln[cys]gln leu val ile gln[cys]glu pro leu glu ala pro glu leu gly thr
GGG CCC CAG[TGT]CAG CTT GTG ATT CAG[TGT]GAG CCT TTG GAG GCC CCA GAG CTG GGT ACC

              210                       220
met asp[cys]thr his pro phe gly[asn phe ser]phe ser gln[cys]ala phe ser[cys]
ATG GAC[TGT]ACT CAC CCC TTT GGA[AAC TTC AGC]TTC AGC TCA CAG[TGT]GCC TTC AGC[TGC]

      230                             240
ser glu gly thr[asn leu thr]gly ile glu glu thr thr[cys]gly pro phe gly[asn trp
TCT GAA GGA ACA[AAC TTA ACT]GGG ATT GAA GAA ACC ACC[TGT]GGA CCA TTT GGA[AAC TGG

      250                             260
ser]ser pro glu pro thr[cys]gln val ile gln[cys]glu pro leu ser ala pro asp leu
TCA]TCT CCA GAA CCA ACC[TGT]CAA GTG ATT CAG[TGT]GAG CCT CTA TCA GCA CCA GAT TTG

      270                             280
gly ile met[asn[cys]ser]his pro leu ala ser phe ser phe thr ser ala[cys]thr phe
GGG ATC ATG[AAC[TGT]AGC]CAT CCC CTG GCC AGC TTC AGC TTT ACC TCT GCA[TGT]ACC TTC

      290                             300
ile[cys]ser glu gly thr glu leu ile gly lys lys lys thr ile[cys]glu ser ser gly
ATC[TGC]TCA GAA GGA ACT GAG TTA ATT GGG AAG AAG AAA ACC ATT[TGT]GAA TCA TCT GGA

      310                             320
ile trp ser[asn pro ser]pro ile[cys]gln lys leu asp lys ser phe ser met ile lys
ATC TGG TCA[AAT CCT AGT]CCA ATA[TGT]CAA AAA TTG GAC AAA AGT TTC TCA ATG ATT AAG

      330                            340       Stop Transfer Sequence
glu gly asp tyr asn pro leu phe ile pro val ala val met val thr ala phe ser gly
GAG GGT GAT TAT AAC CCC CTC TTC ATT CCA GTG GCA GTC ATG GTT ACT GCA TTC TCT GGG
                 ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●

      350                             360
leu ala phe ile ile trp leu ala arg arg leu lys lys gly lys lys ser lys arg ser
TTG GCA TTT ATC ATT TGG CTG GCA AGG AGA TTA AAA AAA GGC AAG AAA TCC AAG AGA AGT
●●●●●●●●●●●●●●●●●●●●●●●●●●●●●

      370      372
met asn asp pro tyr OC
ATG AAT GAC CCA TAT TAA ATCGCCCTTGGTGAAAGAAAATTCTTGGAATACTAAAAATCATGAGATCCTTTA

AATCCTTCCATGAAACGTTTTGTGTGGTGGCACCTCCTACGTCAAACATGAAGTGTGTTCCTTCAGTGCATCTGGCAA

GATTTCTACCCGACCAACAGTTCCTTCAGCTTCCATTTCGCCCCTCATTTATCCCTCAACCCCCAGCCCACAGGTGTT

TATACAGCTCAGCTTTTTGTCTTTTCTGAGGAGAAACAAATAAGACCATAAGGGAAAGGATTCATGTGGAATATAAAG

ATGGCTGACTTTGCTCTTTCTTGACTCTTGTTTTCAGTTTCAATTCAGTGCTGTACTTGATGACAGACACTTCTAAAT

GAAGTGCAAATTTGATACATATGTGAATATGGACTCAGTTTTCTTGCAGATCAAATTTCACGTCGTCTTCTGTATACT

GTGGAGGTACACTCTTATAGAAAGTTCAAAAAGTCTACGCTCTCCTTTCTTTCTAACTCCAGTGAAGTAATGGGGTCC

TGCTCAAGTTGAAAGAGTCCTATTTGCCACTGTAGCCTCGCCGTCTGTGAATTGGACCATCCTATTTAACTGGCTTCAG

GCCTCCCCACCTTCTTCAGCCACCTCTCTTTTTCAGTTGGCTGACTTCCACACCTAGCATCTCATGAGTGCCAAGCAA

AAGGAGAGAAGAGAGAAATAGCCTGCCGCGGTTTTTTTAGTTTGGGGGTTTTGCTGTTTCCTTTTATGAGACCCATTCCT

ATTTCTTATAGTCAATGTTTCTTTTATCACGATATTATTAGTAAGAAAACATCACTGAAATGCTAGCTGCAAGTGACA

TCTCTTTGATGTCATATGGAAGAGTTAAAACAGGTGCGAGAAATTCCTTGATTCACAATGAAATGCTCTCCTTTCCCCT

GCCCCCAGAACTTTTATCCCACTTACCTAGATTCTACATATTCTTTAAATTTCATCTCAGGCCTCCCTCAACCCCACGG

GGCCCCCAGCACACTGGAATTC
```

Fig. 1

GAATTCTCGAGCTCGTCGACCACGCCCTCCTTGTGCAAGAACTCTGAGCCCCAGGTGCAGGAGGCTGAGGCCTGCATAG

```
                              1                        10
                              met val phe pro trp arg [cys] glu gly thr tyr trp gly
AGACTTGCAGAGAGACCCAGCCAAGCC   ATG GTG TTT CCA TGG AGA  TGT  GAG GGT ACT TAC TGG GGC

                              20              Signal Sequence          30
ser arg asn ile leu lys [leu trp val trp thr leu leu [cys][cys] asp phe leu ile] his
TCG ACG AAC ATC CTG AAG  CTC TGG GTC TGG ACA CTG CTC  TGT  TGT  GAC TTC CTG ATA  CAC

            N-Terminus↓      40
his gly thr his [cys] trp thr tyr his tyr ser glu lys pro met asn trp glu asn ala
CAT GGA ACT CAC  TGT  TGG ACT TAC CAT TAT TCT GAA AAG CCC ATG AAC TGG GAA AAT CCT

                              60                              70
arg lys phe [cys] lys gln [asn tyr thr] asp leu val ala ile gln asn lys arg glu ile
AGA AAG TTC  TGC  AAG CAA  AAT TAC ACA  GAT TTA GTC GCC ATA CAA AAC AAG AGA GAA ATT

glu tyr leu glu asn thr     80  leu pro lys ser pro tyr tyr tyr trp ile gly ile arg lys
                                                                90
GAG TAT TTA GAG AAT ACA TTG CCC AAA AGC CCT TAT TAC TAC TGG ATA GGA ATC AGG AAA

                              100                             110
ile gly lys met trp thr trp val gly thr [asn lys thr] leu thr lys glu ala glu asn
ATT GGG AAA ATG TGG ACA TGG GTG GGA ACC  AAC AAA ACT  CTC ACT AAA GAA GCA GAG AAC

                              120                             130
trp gly ala gly glu pro asn asn lys lys ser lys glu asp [cys] val glu ile tyr ile
TGG GGT GCT GGG GAG CCC AAC AAC AAG AAG TCC AAG GAC GAC  TGT  GTG GAG ATC TAT ATC

                              140                             150
lys arg glu arg asp ser gly lys trp asn asp asp ala [cys] his lys arg lys ala ala
AAG AGG GAA CGA GAC TCT GGG AAA TGG AAC GAT GAC GCC  TGT  CAC AAA CGA AAG GCA GCT

                              160                             170
leu [cys] tyr thr ala ser [cys] gln pro gly ser [cys] asn gly arg gly glu [cys] val glu
CTC  TGC  TAC ACA GCC TCT  TGC  CAG CCA GGG TCT  TGC  AAT GGC CGT GGA GAA  TGT  GTG GAA

                              180                             190
thr ile asn [asn his thr] [cys] ile [cys] asp ala gly tyr tyr gly pro gln [cys] gln tyr
ACT ATC AAC  AAT CAC ACG  TGC  ATC  TGT  GAT GCA GGG TAT TAC GGG CCC CAG  TGT  CAG TAT

                              200                             210
val val gln [cys] glu pro leu glu ala pro glu leu gly thr met asp [cys] ile his pro
GTC GTC CAG  TGT  GAG CCT TTG GAG GCC CCT GAG TTG GGT ACC ATG GAC  TGC  ATC CAC CCC

                              220                             230
leu gly [asn phe ser] phe gln ser lys [cys] ala phe [asn [cys] ser] glu gly arg glu leu
TTG GGA  AAC TTC AGC  TTC CAG TCC AAG  TGT  GCT TTC  AAC TGT TCT  GAG GGA AGA GAG CTA

                              240                             250
leu gly thr ala glu thr gln [cys] gly ala ser gly [asn trp ser] ser pro glu pro ile
CTT GGG ACT GCA GAA ACA CAG  TGT  GGA GCA TCT GGA  AAC TGG TCA  TCT CCA GAG CCA ATC

                              260                             270
[cys] gln val val gln [cys] glu pro leu glu ala pro glu leu gly thr met asp [cys] ile
 TGC  CAA GTG GTC CAG  TGT  GAG CCT TTG GAG GCC CCT GAG TTG GGT ACC ATG GAC  TGC  ATC

                              280                             290
his pro leu gly [asn phe ser] phe gln ser lys [cys] ala phe [asn [cys]ser] glu gly arg
CAC CCC TTG GGA  AAC TTC AGC  TTC CAG TCC AAG  TGT  GCT TTC  AAC TGT TCT  GAG GGA AGA

                              300                             310
glu leu leu gly thr ala glu thr gln [cys] gly ala ser gly [asn trp ser] ser pro glu
GAG CTA CTT GGG ACT GCA GAA ACA CAG  TGT  GGA GCA TCT GGA  AAC TGG TCA  TCT CCA GAG

                              320                             330
pro ile [cys] gln glu thr [asn arg ser] phe ser lys ile lys glu gly asp tyr asn [pro]
CCA ATC  TGC  CAA GAG ACA  AAC AGA AGT  TTC TCA AAG ATC AAA GAA GGT GAC TAC AAC  CCC

                              340    Stop Transfer Sequence        350
[leu phe ile pro val ala val met val thr ala phe ser gly leu ala phe leu ile trp]
 CTC TTC ATT CCT GTA GCC GTC ATG GTC ACC GCA TTC TCG GGG CTG GCA TTT CTC ATT TGG

                              360                        370   372
[leu ala] arg arg leu lys lys gly lys lys ser gln glu arg met asp asp pro tyr OP
 TTG GCA  AGG CGG TTA AAA AAA GGC AAG AAA TCT CAA GAA AGG ATG GAT GAT CCA TAC TGA
```
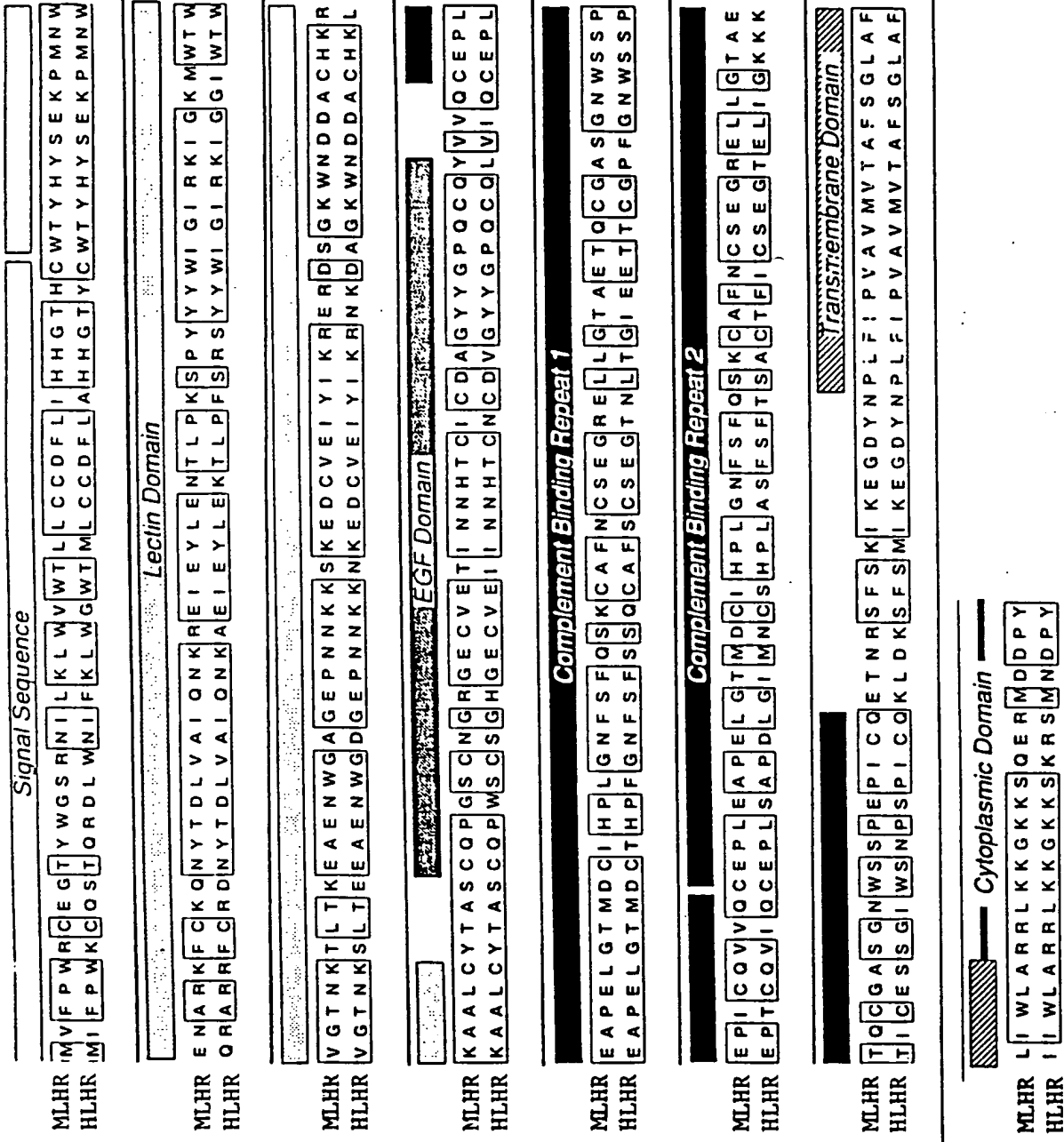
TTCATCCTTTGTGAAAGGAAAGCCATGAAGTGCTAAAGACAAAACATTGGAAAATAACGTCAAGTCCTCCCGTGAAGA

TTTTACACGCAGGCATCTCCCACATTAGAGATGCAGTGTTTGCTCAACGAATCTGCAAGGGATTTCTTCATGACCAACA

GCTCCTCCTAATTTCCCCTCGCTCATTCATCCCATTAACCCTATCCCATAATGTGTGTCTATACAGAGTAGTATTTTA

TCATCTTTTCTGTGGAGGAACAAGCAAAAGTGTTACTGTAGAATATAAAGACAGCTGCTTTTACTCTTTCCTAACTCT

TGTTTCCTAGTTCAATTCAGCCACAGAAGCTAATGCCAAACACACTGAAAATATGATCCATGACTAATTGCAAACTCAG

ACTCCTTGCGCCATAGTACGTACCCTATGTAACATCGACAAAAATCTTTCATTTCCACCTCCAAAGAACAGTGCTCTAT

TCAAGTTGGCAAAGTCCTACTTCCTCTGTAGACCCACTATCTGTGAGTGACAGCCACTGTAGCCTGTTCACATTAACCT

TCCCCATCTCCTTTTCCTAGGAGAATAATTCCACACACTGCACCCCATGATGGCCACCAAACATCAAAGAAGGGAAAA

TCTCCTGCCATTGAGTTTTAGTTTTGAGTTTTCCCTTCTCTTTATTACATCTCTCATGGTTCCTTGAAGTCACTGTTCT

GATGATTATTAATAGTTAATGATAACACAACCCACTCTCTTGGAGCTGATGTTATGAAGACAACACGTAGAAAAATTC

CTGGGCCTCAGGCTCGGAGTCGACACCCTTTTCTTTCCCTAACATCTTCTACTCAGATACCTAAATTTAAGATTCAGGACA

GCTCGTCCCCAACTCTTACCATGTCTTTTATAACTTGCCTCCTTAACTTGCCCCAACCCTGTAGCCTATCTCATTTTCTCGC

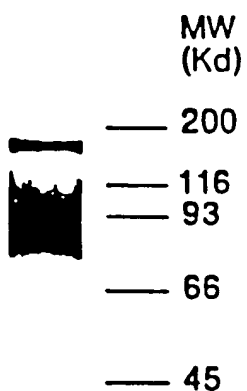TTCACTCTGCAAGGTTTATAACATGATGAATTTAAATACAAAAAAAAAAAAAAAA

Fig. 2

Signal Sequence

MLHR M VFPWRC EGT YWGSRNI LKLWWT L LCCDFLI HHGT HCWTYHYSEKPMNW
HLHR MI FPWK CQS TQRDLWNI FKLWGWT ML CCDFLA HHGT YCWTYHYSEKPMNW

Lectin Domain

MLHR ENARKF CKQNYTDLVAIQNK REIEYLE NTLPKS PYYYWIGIRKIGKMWTW
HLHR QRARRF CRDNYTDLVAIONKA EIEYLEK TLPFSR SYYWIGIRKIGGIWTW

MLHR VGTNKTLT KEAENWGAGEPNNKKSKEDCVEIYIKRERDSGKWNDDACHKR
HLHR VGTNKSLTEEAENWGDGEPNNKKNKEDCVEIYIKRNKDAGKWNDDACHKL

EGF Domain

MLHR KAALCYTASCQPGSCNRGECVETINNHTCICDAGYYGPQCQYVQCEPL
HLHR KAALCYTASCQPWSCSGHGECVEIINNHTCNCDVGYYGPQCQLVIQCEPL

Complement Binding Repeat 1

MLHR EAPELGTMDCIHPLGNFSFQSKCAFNCSEGRELLGTAETQCGASGNWSSP
HLHR EAPELGTMDCTHPFGNFSFSSQCAFSCSEGTNLTGIEETTCGPFGNWSSP

Complement Binding Repeat 2

MLHR EPICQVVQCEPLEAPELGTMDCIHPLGNFSFQSKCAFNCSEGRELLGTAE
HLHR EPTCQVIQCEPLSAPDLGIMNCSHPLASFSFTSACTFICSEGTELIGKKK

MLHR TQCGASGNWSSPEPICQETNRSFSKIKEGDYNPLFIPVAVMVTAFSGLAF
HLHR TICESSGIWSNPSPICQKLDKSFSMIKEGDYNPLFIPVAVMVTAFSGLAF

Transmembrane Domain

Cytoplasmic Domain

MLHR LIWLARRLKKGKKKSQERMDDPY
HLHR IIWLARRLKKGKKKSKRSMNDPY

MW
(Kd)

— 200

— 116
— 93

— 66

— 45

Fig. 4 A

Fig. 4 B

```
T H    K M        KF K        VVIL  K
1          10        20          30
XTYHYSEKPMNWENARKFXKQNYTDLVAIQNKXXIEYL
```

Fig. 4 c

```
      A    A            C        A   C
5' GAG AAG CCC ATG AAT TGG GAG AAT GC  3'
```

Fig. 5

Fig. 6A

Drickamer
MLHR
Hu.HepLec
Barn.Lec
Ra.HepLec
Ch.HepLec
Hu.IgERec
RaHepLec2
Ra.ASGRec
Ra.IRP
Ra.MBP
Ra.MBDA
Ra.KCBP
FlyLec
Rab.Surf
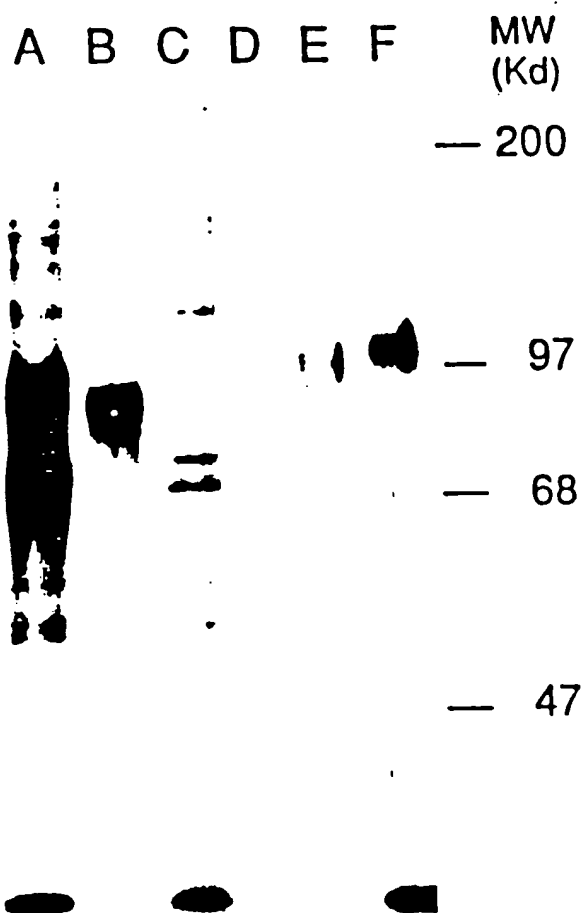
Fig. 6B

MLHR	160	C	QPGSCNGRG	ECVET	INNHTCI	CDAGYYGPQ	CQY
Notch	1021	C	TESSCLNGGS	CIDGL	NGGYNCS	CLAGYSGANC	QY
S.purp.	61	C	ASAPCQNGGV	CIDGV	NGYMCD	CQPGYTGTH	CET
Pro.Z	1	C	ASQPCLNNGS	CQDSI	RGYACT	CAPGYEGPN	CAF
Fact.X	61	C	EGHPCLNQGH	CKDGI	GDYTCT	CAEGFEGKN	CEF
Fact.VII	61	C	ETSPCQNQGK	CKDGL	GEYTCL	EGFEGKNC	EL
Fact.IX	1	C	ESNPCLNGGM	CKDDI	NSYECW	CQAGFEGTN	CEL
Lin-12	361	C	LENPCSNGGV	CHQHRESFS	CDCPPGFYGN	CEQ
Fact.XII	121	C	RTNPCLHGGR	CLEVEGHRL	CHCPVGYTGPF	CDV
Mu.egf	841	C	GPGGCGSHARCVSDGETAE	CQCLKGFARDGNL	C

Row labels (repeated for each alignment block):

MLHR
HuComH
MuComH
HuBeta
HuCR1
EBV/C3d
HuC2
HuB
MuC4b
HuC1s
HuC4b
HuDAF
VacSecP

Fig. 7

PROTEIN MOTIFS IN THE LYMPHOCYTE HOMING RECEPTOR